INTI JOURNAL | eISSN:2600-7320
Vol.2020:59

Forecasting Stock Price using ARMA Model

Koh Wei Sin[1*], Heng Hong Sheng[1], Wong Chong Zhi[1], Lai Pui Ling[1], Charisma Dass[1]

[1]Faculty of Business, Communication & Law, INTI International University
Persiaran Perdana BBN, Putra Nilai, 71800 Nilai, Negeri Sembilan, Malaysia.

[*]**Email:** weisin.koh@newinti.edu.my

Abstract

Forecasting is the process of making predictions based on the historical data. In this paper, we took the daily opening stock prices of Maxis Berhad from Jan 2010 to Dec 2017 to analyze and forecast the opening stock prices from Jan 2018 to Dec 2019. Before the modelling part, we examined the stationarity of the time series data. The data were found to be non-stationary and some transformation procedures were implemented onto the data such as differencing and log transformations. After that, the transformed data were modeled with Autoregressive Moving Average (ARMA) models through Eviews software. ARMA model is the combination of AR(p) and MA(q) models. In this study, we examined ARMA models of order p+q up to 5 order. Then, we did the Global and Coefficients tests to produce the selected models. The selected models will then be inspected based on standard error, r squared and some criteria to obtain the best model. The best model is used to derive the predicted time series data. The predicted time series data is then detransformed and compared with the real daily opening stock prices of Maxis Berhad from Jan 2018 to Dec 2019. Finally, the predicted daily opening stock prices were shown to be having high accuracy with the Mean Absolute Percentage Error (MAPE) of 1.41%.

Keywords

Forecasting, ARMA model, Time Series

Introduction

Stock price is a time series data. The change of the stock price is depending on the volatility of the stock price. Higher volatility means more drastic change of the stock price. In statistics, volatility of the stock price can be measured through the standard deviation and variance of the stock price.

International Conference on Innovation and Technopreneurship 2020
Submission: 4 September 2020; Acceptance: 7 October 2020

Copyright: © 2020. All the authors listed in this paper. The distribution, reproduction, and any other usage of the content of this paper is permitted, with credit given to all the author(s) and copyright owner(s) in accordance to common academic practice. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license, as stated in the website: https://creativecommons.org/licenses/by/4.0/

Forecasting a stock price is about developing a mathematical model to predict the future value of the stock price and its trend. By predicting the future value of the stock allows the investor to make better decision whether to buy, sell or hold the stock.

An Autoregressive and Moving Average (ARMA) model is the combination of autoregressive AR (p) and moving average MA(q) models. ARMA model has been used to forecast stock prices (Anaghi & Norouzi, 2012; Mondal, Shit, & Goswami, 2014), wind speed and direction (Erdem & Shi, 2011), electricity demand load (Pappas et al., 2010), household electric consumption (Chujai et al., 2013), aluminium price (Ru & Ren, 2012) and other time series data. The general form of ARMA model is

$$\hat{Y}_t = \theta + \alpha_1 Y_{t-1} + \cdots + \alpha_p Y_{t-p} + \beta_0 u_t + \beta_1 u_{t-1} + \cdots + \beta_q u_{t-q} \quad (1)$$

where, Y_t is the value at time t, θ is a constant term, α_i and $\beta_j$are the coefficients and u_t is the white noise stochastic error term (Gujarati, Porter, & Gunasekar, 2017).

In order to apply the ARMA model, the stationary must be assumed (Gujarati et al., 2017). A time series data is known as stationary if mean and variance are constant over time . (Gujarati et al., 2017; Brockwell & Davis, 2016) Normally, the data will be non stationary and the time series data need to be transformed to achive stationary.

In this paper, we will forecast the daily opening stock prices of Maxis Berhad. Ten years daily stock prices of Maxis Berhad data from year 2010 to 2019 will be taken from Yahoo Finance. The time series data of the daily opening stock price from Jan 2010 to Dec 2017 are analyzed to obtain the best ARMA model. And the remaining daily opening stock price from Jan 2018 to Dec 2019 will be used to evaluate the forecasting results.

Methodology

Before the modelling part, the assumption of stationary process need to be justified. ANOVA and Levene's test were implemented by using IBM SPSS version 26.0 to test the equality of mean and variance at different time periods (Hinton, McMurray, & Brownlow, 2014). When mean and variance showed changes over the time, transformation such as differencing and log function need to be done onto the data to obtain stationary.

After the stationary of the time series data is achieved through transformations of the data, ARMA models with $p+q \leq 5$ were formulated. There are altogether 20 possible ARMA models being developed. All the ARMA models and other numerical results were estimated using the Eviews version 10.0.

Then, coefficient and global tests were conducted using the numerical results. The purpose of the global and coefficient tests are to test whether the coefficients equal to zero, where global test observe the F-value while coefficient tests check the coefficient individually through the t-value (Gujarati et al., 2017). The null hypothesis for global test is all the cofficients are equal to zero while coefficient test is that particular individual coefficient equal to zero. Coefficients which tested to be significantly zero do not contribute to the model and are eliminated. The selected models from the elimination process were then be inspected based on the standard error, r squared and some criteria to determine the best model.

The best ARMA model is used to derive the predicted time series data. The predicted time series data is then detransformed and compared with the real daily opening stock prices of Maxis Berhad from Jan 2018 to Dec 2019 in Microsoft Excel. Finally, the accuracy is observed by taking the Mean Absolute Percentage Error (MAPE).

Results and Discussion

Firstly, we check the stationary of the daily opening stock price from Jan 2010 to Dec 2017. The data is divided into five groups to perform the equality of mean and variance tests at 5% of significance level. Table 1 is the ANOVA table for the equality of mean test where the null hypothesis is all means are equal between the groups. And Table 2 is the Levene's test output for the equality of variance test where the null hypothesis is all variances are equal between the groups. Since the p-value is less than 0.05, null hypotheses are rejected. In other words, the means and the variance are not equal over the time period. Hence, the daily opening stock price is non-stationary as shown also in Figure 1.

Table 1. ANOVA table of the opening stock price in five different time period

ANOVA

y

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	528.430	4	132.107	1076.382	.000
Within Groups	242.397	1975	.123		
Total	770.827	1979			

Table 2. Levene's test of the opening stock price in five different time period

Test of Homogeneity of Variances

		Levene Statistic	df1	df2	Sig.
y	Based on Mean	256.700	4	1975	.000
	Based on Median	240.772	4	1975	.000
	Based on Median and with adjusted df	240.772	4	1191.903	.000
	Based on trimmed mean	254.776	4	1975	.000

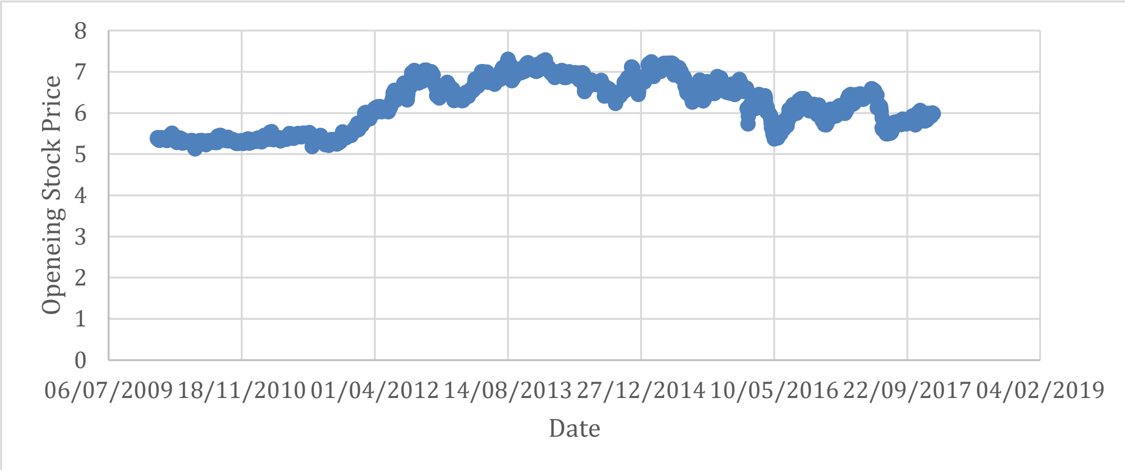

Figure 1. Scatterplot of the Daily Opening Price from Jan 2010 to Dec 2017

Since the time series data is non-stationary, log-transformation was applied onto the data. However, it is still not-stationary as seen in the Figure 2. Then, the first differencing transformation is implemented onto the transformed time series data. After the first differencing, the equality of mean and variance were tested based on the Table 3 and 4. The null hypothesis of the equality of mean test is accepted but rejected for equality of variance. It means the mean is equal over the time period but not the variance. However, the scatterplot in Figure 3 looks stationary with some outliers. We did second differencing transformation but the results are still the same. Hence, the transformation stopped at first differencing of log-transformation to avoid over-differencing (Greunen et al., 2014).

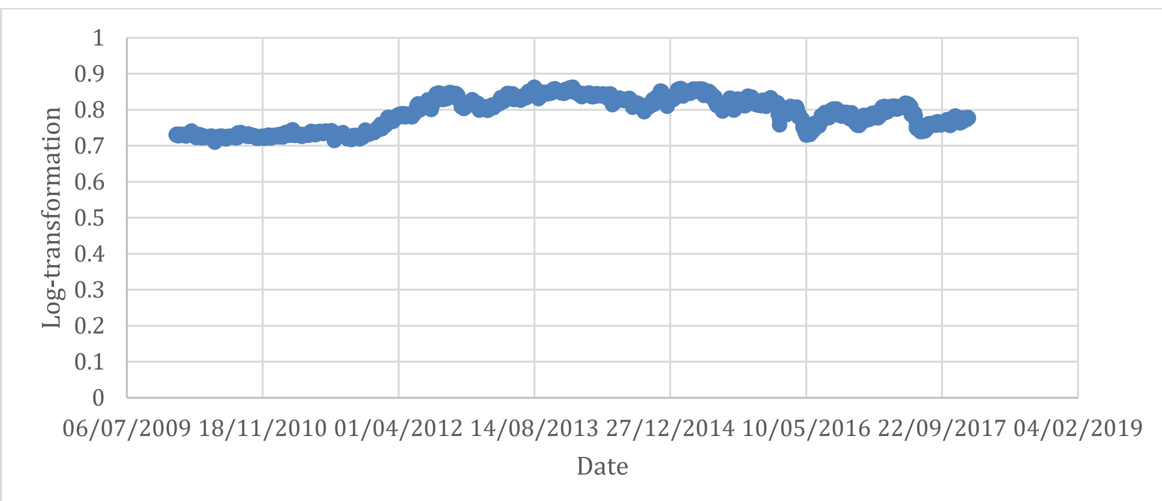

Figure 2. Scatterplot of the log-transformation from Jan 2010 to Dec 2017

Table 3. ANOVA table of the first differencing of log-transformation

ANOVA

1st Differencing of Log

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	.000	4	.000	.575	.681
Within Groups	.032	1974	.000		
Total	.032	1978			

Table 4. Levene's test of the first differencing of log-transformation

Test of Homogeneity of Variances

		Levene Statistic	df1	df2	Sig.
1st Differencing of Log	Based on Mean	43.856	4	1974	.000
	Based on Median	43.427	4	1974	.000
	Based on Median and with adjusted df	43.427	4	1533.970	.000
	Based on trimmed mean	43.908	4	1974	.000

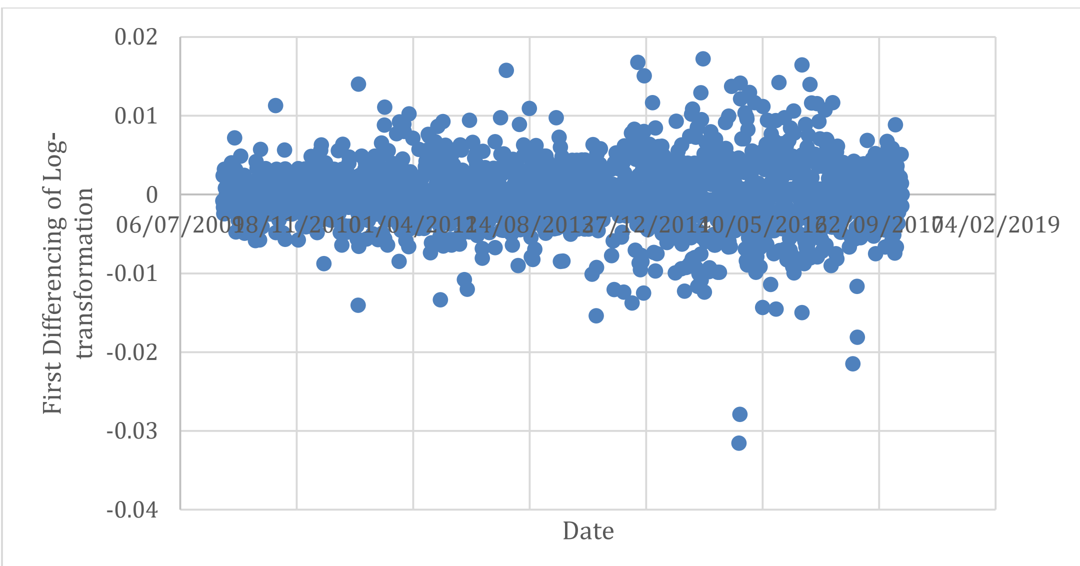

Figure 3. Scatterplot of the first differencing of log-transformation

The first differencing of log-transformation time series data is used to estimate the ARMA models using Eviews 10.0. Then, the global and coefficient tests were run onto the models and elimination of non-significant coefficients is done. Table 5 shows the elimination process and the selected models. Some models after the elimination yield new models while some return to existing models.

Table 5. Summary table of the elimination process through global and coefficient test

Models	**Process**	**Selected**
M1	M1	M1
M2	M2	Eliminated
M3	M3	M3
M4	M4	M4
M5	M5	M5
M6	M6	M6
M7	M7 -> M7.1 = M3	M3
M8	M8 -> M8.1	M8.1
M9	M9 -> M9.1 = M5	M5
M10	M10	M10
M11	M11 -> M11.1	M11.1
M12	M12 -> M12.1 -> M12.2 = M3	M3
M13	M13	M13
M14	M14	M14
M15	M15	M15
M16	M16 -> M16.1 -> M16.2 = M6	M6
M17	M17 -> M17.1 -> M17.2 -> M17.3 = M5	M5
M18	M18 -> M18.1 -> M18.2	M18.2
M19	M19 -> M19.1 = M14	M14
M20	M20 -> M20.1 = M14	M14

The selected models are compared using some criteria such as Akaike Info Criterion (AIC), Hannan-Quinn criter (HQ), Schwarz criterion, R-squared, adjusted R-squared and standard error as shown in Table 6. The highest values for R-squared and adjusted R-squared, and lowest values of other criterions were highlighted. Model M18.2 is chosen to be the best model because it fulfils the most criterions. The model M18.2 consists of AR(2), MA(1) and MA(3), and can be derived as

$$\hat{V}_t = -1.1\mathrm{E}-7 + 0.139387\, V_{t-2} + 1.147496\, u_{t-1} - 0.147531 u_{t-2} \qquad (2)$$

where V_t is the first differencing of log-transformation at time *t*.

The predicted values are computed and detransformed to form the forecasted values. The forecasted values are compared with the daily opening stock price from Jan 2018 to Dec 2019. In Figure 4, we can observe how closely are the forecasted values and the actual daily opening stock price. To evaluate its accuracy, the MAPE is calculated and obtain 1.41%, which is highly accurate with less percentage of error.

Table 6. Comparison of the selected models to obtain the best model

Selected Models	n	k	SSE	SIGMASQ	AIC	HQ	SCHWARZ	SE	R^2	Adjusted R^2
M1	1521	2	0.038	2.50E-05	-7.641	-7.637	-7.63029	0.005	0.34	0.339
M3	1521	3	0.035	2.30E-05	-7.678	-7.673	-7.66409	0.0048	0.392	0.391
M4	1521	3	0.0304	2.00E-05	-7.7	-7.695	-7.68591	0.0045	0.471	0.47
M5	1521	3	0.0312	2.05E-05	-7.699	-7.694	-7.6854	0.0045	0.458	0.457
M6	1521	4	0.0318	2.09E-05	-7.707	-7.701	-7.6897	0.0046	0.447	0.446
M8.1	1521	3	0.0313	2.06E-05	-7.701	-7.696	-7.68693	0.0045	0.456	0.455
M10	1889	5	0.0358	1.89E-05	-7.981	-7.974	-7.96323	0.0044	0.491	0.49
M11.1	1889	4	0.0297	1.57E-05	-8.096	-8.09	-8.08121	0.004	0.577	0.576
M13	1889	5	0.0298	1.58E-05	-8.091	-8.084	-8.07301	0.004	0.576	0.575
M14	1889	5	0.0298	1.58E-05	-8.095	-8.088	-8.07701	0.004	0.576	0.575
M15	1889	6	0.0343	1.81E-05	-8.015	-8.007	-7.9942	0.0043	0.512	0.511
M18.2	1978	4	0.0311	1.57E-05	-8.214	-8.209	-8.19978	0.004	0.58	0.579

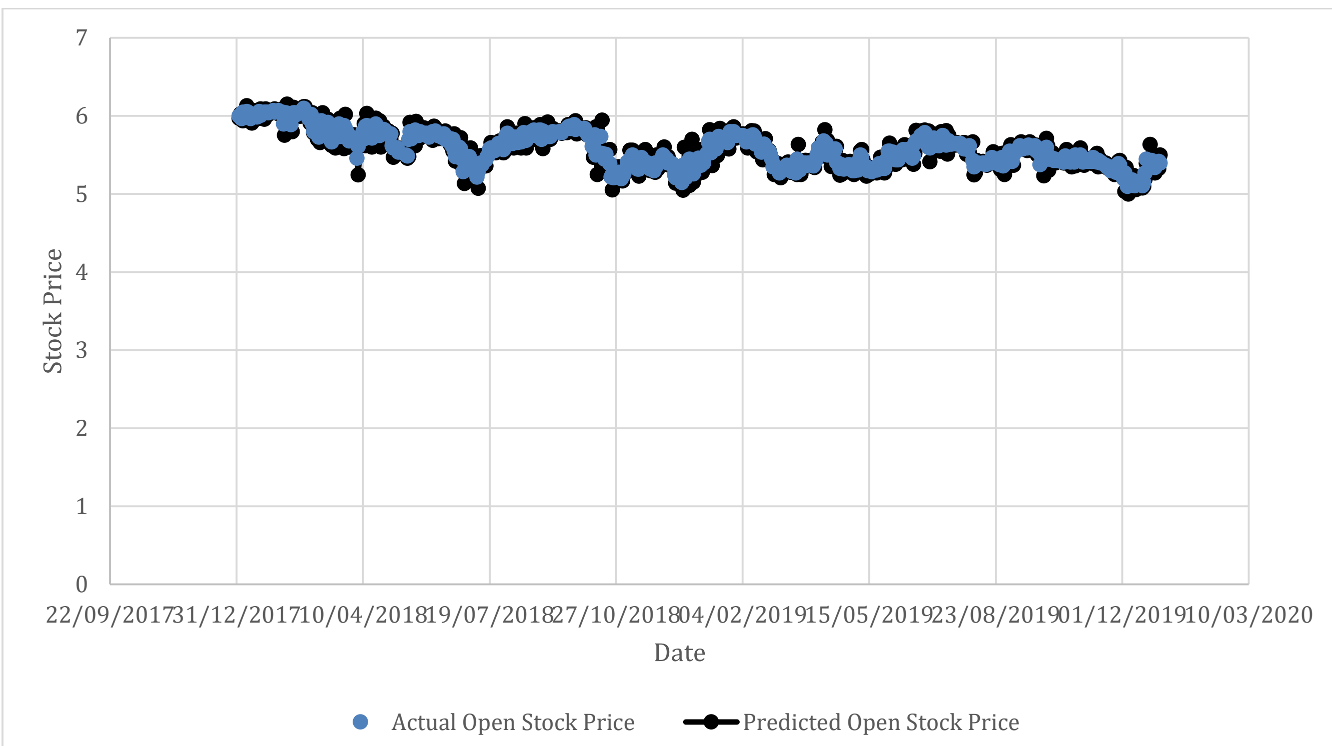

Figure 4. Scatterplot of the comparison of predicted and actual opening Maxis stock price

Conclusions

An ARMA model in equation (2) and some transformations have been developed to forecast the daily opening stock price of Maxis Berhad. This model has been tested for two years forecasting from Jan 2018 to Dec 2019. Eventually, it has been proven to be highly accurate with MAPE 1.41% only, which means less percentage of mean absolute error.

For further study, we can use this ARMA estimation modelling with other stock price. We can also increase the number of possible models to $p+q \leq 10$ to study much more models to derive the best model.

References

Anaghi, M., & Norouzi, Y. (2012). A model for stock price forecasting based on ARMA systems. *2012 2nd International Conference on Advances in Computational Tools for Engineering Applications (ACTEA)* (pp. 265-268). Beirut: IEEE.

Brockwell, P., & Davis, R. (2016). *Introduction to Time Series and Forecasting (3rd ed.).* Springer.

Chujai, P. K. (2013). Time series analysis of household electric consumption with ARIMA and ARMA models. *Proceedings of the International MultiConference of Engineers and Computer Scientists,* (pp. 295-300). Hong Kong.

Erdem, E., & Shi, J. (2011). ARMA based approaches for forecasting the tuple of wind speed and direction. *Applied Energy*, 1405-1414.

Greunen, J., Heymans, A., Heerden, C., & Vuuren, G. (2014). The prominence of stationarity in time series forecasting. *Journal for Studies in Economics and Econometrics*, 1-16.

Gujarati , D., Porter, D., & Gunasekar, S. (2017). *Basic Econometrics.* McGraw-Hill Education.

Hinton, P., McMurray, I., & Brownlow, C. (2014). *SPSS Explained.* Routledge.

Maharaj, E. A. (1996). A significance test for classifying ARMA models. *Journal of Statistical Computation and Simulation*, *54*(4), 305-331.

Mondal, P., Shit , L., & Goswami, S. (2014). Study of effectiveness of time series modeling (ARIMA) in forecasting stock prices. *nternational Journal of Computer Science, Engineering and Applications*, 13-29.

Pappas, S. S. (2010). Electricity demand load forecasting of the Hellenic power system using an ARMA model. *Electric Power Systems Research*, 256-264.

Ru, Y. &. (2012). Application of ARMA model in forecasting aluminum price. *Applied mechanics and materials*, 66-71.